ELSEVIER

Cancer Genetics and Cytogenetics 168 (2006) 89–97

Cancer Genetics and Cytogenetics

Lead article

Linking the human cytogenetic map with nucleotide sequence: the CCAP clone set

Wonhee Jang[a,1], Raluca Yonescu[b,1], Turid Knutsen[b], Theresa Brown[b], Tricia Reppert[b], Karl Sirotkin[a], Gregory D. Schuler[a], Thomas Ried[b,*], Ilan R. Kirsch[b,2]

[a]*National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD*
[b]*Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD*

Received 29 December 2005; accepted 3 January 2006

Abstract We present the completed dataset and clone repository of the Cancer Chromosome Aberration Project (CCAP), an initiative developed and funded through the intramural program of the U.S. National Cancer Institute, to provide seamless linkage of human cytogenetic markers with the primary nucleotide sequence of the human genome. Spaced at 1–2 Mb intervals across the human genome, 1,339 bacterial artificial chromosome (BAC) clones have been localized to chromosomal bands through high-resolution fluorescence in situ hybridization (FISH) mapping. Of these clones, 99.8% can be positioned on the primary human genome sequence and 95% are placed at or close to their precise nucleotide starts and stops. This dataset can be studied and manipulated within generally available public Web sites. The clones are available from a commercial repository. The CCAP BAC clone set provides anchors for the interrogation of gene and sequence involvement in oncogenic and developmental disorders when the starting point is the recognition of a structural, numerical, or interstitial chromosomal aberration. This dataset also provides a current view of the quality and coherence of the available genome sequence and insight into the nucleotide and three-dimensional structures that manifest as Giemsa light and dark chromosomal banding patterns.

1. Introduction

The Cancer Chromosome Aberration Project (CCAP) arose within the context of the Cancer Genome Anatomy Project (CGAP), which is an attempt by the U.S. National Cancer Institute to ‘‘create a catalog of the genes associated with cancer, and to develop technological tools to support the analysis of the molecular profiles of cancer cells and their normal counterparts’’ [1,2]. CCAP [3,4] was based on the realization that cancer is a genetic disease caused by genetic instability; every cancer that has been studied in requisite detail has been shown to contain a genome distinct from the genome of the normal cells from which it arose. Over the past 30 years, it has become apparent that the vast majority if not all of human cancers contain chromosomal aberrations, which may include deletions, insertions, amplifications, and chromosomal translocations. These numerical and structural genomic changes tend to be more numerous in malignant tumors than in benign ones [5,6] and are often associated with distinctive prognostic, clinical, and histopathological features [7–10].

The identification of *MYC* (alias c-*MYC*) dysregulation by immunoglobulin enhancers, which was accomplished via the cloning and characterization of the chromosomal translocation specifically associated with the development of Burkitt lymphoma [11–13], established as a precedent the utility of exploring cancer-associated chromosomal aberrations as a pathway for studying the mechanisms of malignant transformation. This exploration would be greatly facilitated if there were a streamlined process to expedite the task of relating a particular aberration to the genes and sequences affected by it. The goal of CCAP was to aid in this streamlining by developing a set of tools that could seamlessly link cytogenetic locations with primary human genome sequence. Here we describe such a tool set.

[1] Contributed equally to this work.
[2] Current address: Oncology Research, Amgen, 1201 Amgen Court West, AW1-J4144, Seattle, WA 98119-3105.
[*] Corresponding author. Tel.: (301) 594-3118; fax: (301) 435-4428.
E-mail address: RiedT@mail.nih.gov (T. Ried).

0165-4608/06/$ – see front matter
doi:10.1016/j.cancergencyto.2006.01.001

2. Materials and methods

2.1. Clone selection for the CCAP set

Initially, bacterial artificial chromosome (BAC) clones were sought from investigators furthest along in the development of contig maps for specific human chromosomes. As the Human Genome Project [14] continued, BAC clones were specifically selected based on their predicted coverage and spacing on a given chromosome, but even more because of their inclusion in the queue of clones scheduled to be sequenced in their entirety as part of the Human Genome Project.

BAC clones were obtained from the following investigators and centers: chromosomes 1, 5, 18: Pieter de Jong and Norma Nowak, Roswell Park Cancer Institute, Buffalo, NY; chromosomes 2, 3, 4, 6, 8, 9, 10, 11, 13, 19, 20, 21, Y: Pieter de Jong, Children's Hospital of Oakland Research Institute, Oakland, CA; chromosome 7: Eric Green, National Human Genome Research Institute, Bethesda, MD; chromosome 12: Raju Kucherlapati, Dana Farber Cancer Institute, Boston, MA; chromosomes 14, 15, 16, 17, X: Vivien Cheung, University of Pennsylvania, Philadelphia, PA; and chromosome 22: Nigel Carter, Sanger Centre, Cambridge, England, UK.

For chromosomes 2, 4, 6, 8, 9, 10, 11, 13, 19, and Y, potential clones for fluorescence in situ hybridization (FISH) mapping were selected based on the draft human genome assembly available as of February 2001. Selection criteria were that (i) it had to be a BAC clone from the RP11 library, (ii) it had to contain either high-throughput genome (HTG) sequence or have one BAC-end sequence, and (iii) the spacing of the clones had to be 1–2 Mb on the draft genome.

2.2. FISH mapping

The BAC clones were obtained as glycerol stocks or agar stabs from the various investigators. DNA was purified as per published protocol (Autogen, Framingham, MA). The purified BAC DNA was used for high-resolution, dual-color FISH mapping [15] onto prometaphase chromosomes that were prepared following a standard protocol [16]. All CCAP clones were sent to Open Biosystems (Huntsville, AL) as purified glycerol stocks or agar stabs.

2.3. BAC-end sequencing

Clones were sent to The Institute for Genome Research (TIGR, Gaithersburg, MD) or to the Core Genotyping Facility at the National Cancer Institute for BAC-end sequencing if both end sequences were not available in the GenBank genetic sequence database and if end information could not be obtained from the insert sequences. All BAC-end sequences produced were deposited in the GenBank.

2.4. Identifying sequence anchors for clone placement on the human genome

Three types of sequences that are associated with clones were identified and used for clone placement on the human genome. Identifying these sequences relies solely on data in GenBank where submitters have provided a clone name in the sequence submission.

2.4.1. BAC-end sequences

The BAC-end sequences of CCAP clones were obtained from the GSS division of GenBank [17,18], masked of low-complexity regions and interspersed repeats with the RepeatMasker, and aligned to the human genome using the Megablast program [19]. For BAC-end sequences that failed to align to the genome after masking, alignment was performed without repeat masking. Alignments of >100 bp and $>90\%$ identity were retained, and the longest and best match was selected. If a BAC-end was repetitive and was aligned to multiple regions on the genome, the alignment compatible with the mate-pair end (the BAC-end sequenced from the opposite end of the same clone) was chosen.

2.4.2. Insert sequences

The insert sequences of CCAP BAC clones were retrieved from the GenBank PRI division. These insert sequences can be a finished sequence, high-throughput genome sequence (HTGS) phase 3, or unfinished sequence, HTGS phase 1 or 2 (see http://www.ncbi.nlm.nih.gov/genome/guide/glossary.htm for definition of HTGS phase 1, 2, and 3). In brief, HTGS phase 3 sequences have no gaps, and their sequences are of high quality, usually <1 sequencing error in 10 kb, but they may or may not cover the entire clone insert. HTGS phase 2 sequences have gaps and fragments within the clone but are ordered and oriented. HTGS phase 1 sequences may have gaps and fragments within the clone and are not ordered and oriented.

Since we were interested in marking the extent of a clone on the genome, we included previous versions of HTG sequences in the analysis if they were in a different phase than the most recent sequence, because HTG sequences in previous versions may contain more sequences than the more recent version. We also tried to identify clone ends from HTG sequences. This was done by aligning vector sequences, checking annotations explicitly marking clone ends, or having restriction enzyme recognition sites at the start and end of HTG sequences. HTG sequences that contain clone end information were also used for analysis. HTG sequences associated with the CCAP clones were aligned to genomic contigs using the Megablast program; alignments with $>98\%$ identity and length of >100 bp were retained. Afterwards, each HTG sequence was placed onto the highest scoring contig and spurious hits were removed. Previous versions of HTG sequences were used

only if their placements overlapped with that of the most recent HTG sequence.

2.4.3. STS

The sequence tagged site (STS) content of a clone was determined by a combination of hybridization and polymerase chain reaction (PCR) [20]. STSs were located on the genome using an electronic PCR software tool [21].

After all three types of sequence anchors were placed on the genomic contigs, clone placement was finalized based on consensus of placements of all sequence anchors along with FISH mapping data. If a single sequence anchor was placed to a discordant region of the genome compared to FISH data, but multiple other sequence anchors were concordant, the discordant placement of a single sequence anchor was ignored. If there was more than one independent piece of sequence data assigning a clone to two separate distinct locations, each placement was retained and a notation was made (this occurred for 11 of 1,339 clones).

3. Results

3.1. FISH mapping of BAC clones

BAC clones were used in pairs and were hybridized onto high-resolution prometaphase chromosomes using dual-color FISH. Because we wished to be able to orient more than one clone within a given band, we arbitrarily divided an 850-band ideogram [22] into approximately 1 Mb segments based on the crude estimate of the size of the human genome as 3,000 Mb. We chose to add a lowercase letter (beginning with 'a') to the ISCN 1995 [22] nomenclature to indicate these approximated subregions within each band. This resulted in 3,207 chromosomal subregions. For example, the large band 1p31.1 was further subdivided into 11 subregions (1p31.1a, 1p31.1b, …, 1p31.1k, centromeric to telomeric) and we ordered five BAC clones in this band by using this a-b-c notation. This is not intended to be a new nomenclature, but only to allow for the orientation, centromeric to telomeric, of CCAP BAC clones that all localize to the same ISCN-described cytogenetic band.

Figure 1 provides an example of this high-resolution, dual-color FISH mapping. One of us (R.Y.) was responsible for the majority of all the high-resolution mapping and, furthermore, monitored and reviewed the mapping efforts of others and was the final arbiter of clone placement. Each BAC clone included in the CCAP set is required to hybridize to only one site in the genome. Each clone also was required to have sufficient sequence information to allow for its placement on the human genome.

3.2. Clone placement on the genome

Placing these clones on the human genome with precise clone boundaries enables the investigation of the sequenced genome starting from cytogenetic landmarks. This was accomplished by placing the clone using its available sequence onto the genome.

The extent of available sequence for the CCAP clones is shown in Table 1. Among these sequences, 96% (2,638 of 2,761) of BAC-end sequences, 99% (917 of 927) of high-throughput genome (HTG) sequences, and 95% (942 of 990) of STSs from the CCAP set were placed on the genome. Using the sequence placement data, we were able to place 99% (1,325 of 1,339) of the CCAP clones at a unique location. For 0.8% (11 of 1,339), localization could only be narrowed down to either of two locations because placement of their sequences could not be resolved to a single location. A final 0.2% (3 of 1,339) of the clones could not be placed on the human genome at all, because sequences of these clones failed to align to the genome.

The determination of precise clone boundaries serves to focus attention of investigators of a cancer-associated chromosomal aberration on the specific genes and other sequences in a bounded region. This information facilitates the elucidation of both the causes and consequences of the chromosomal aberration—the goal of such investigations. The degree of precision in the determination of clone boundaries on the genome depends on the kind of sequences used for clone placement.

Figure 2 summarizes the types of sequences used for determining clone boundaries on the genome. Clones in categories A–D, where the clone boundary is determined by BAC-ends or by complete insert sequences, have precisely determined clone boundaries at or within a few dozen bases of the actual insert end. For clones in categories E and F, where the clone is placed by insert sequences, it is reasonable to assume that the placement represents the extent of the clones as well. In categories A–F combined, 94.9% of CCAP clones have quite well-defined clone boundaries on the human genome, making it possible to integrate cytogenetic data with the human genomic sequences with, essentially, nucleotide resolution. For 2.9% of the CCAP clones, those in category G, one boundary is known but the other end is not known. For 1.1% of clones, categories H and I, clone boundaries are only approximate to within 150 kb.

Direct coverage of the genome by CCAP clones is approximately 223 Mb, which comprises 7.8% of the assembled human genome in the NCBI build 35. Average clone spacing is 1.9 Mb excluding centromeric and telomeric parts of the genome.

3.3. Online resources and the clone distributor

All CCAP data can be viewed in the CCAP Web site (http://cgap.nci.nih.gov/Chromosomes/CCAP_BAC_Clones). The site lists each clone name, FISH-mapping result, all associated sequences, and genome placement results. The rare sequences whose genomic positions disagree with their clone positions are marked, to distinguish them from concordant sequences. An example of the data representation is shown in Fig. 3.

The NCBI MapViewer (see Appendix for this and other exact Internet URL addresses) also has two tracks for CCAP clones: one is called NCI_clone and it displays

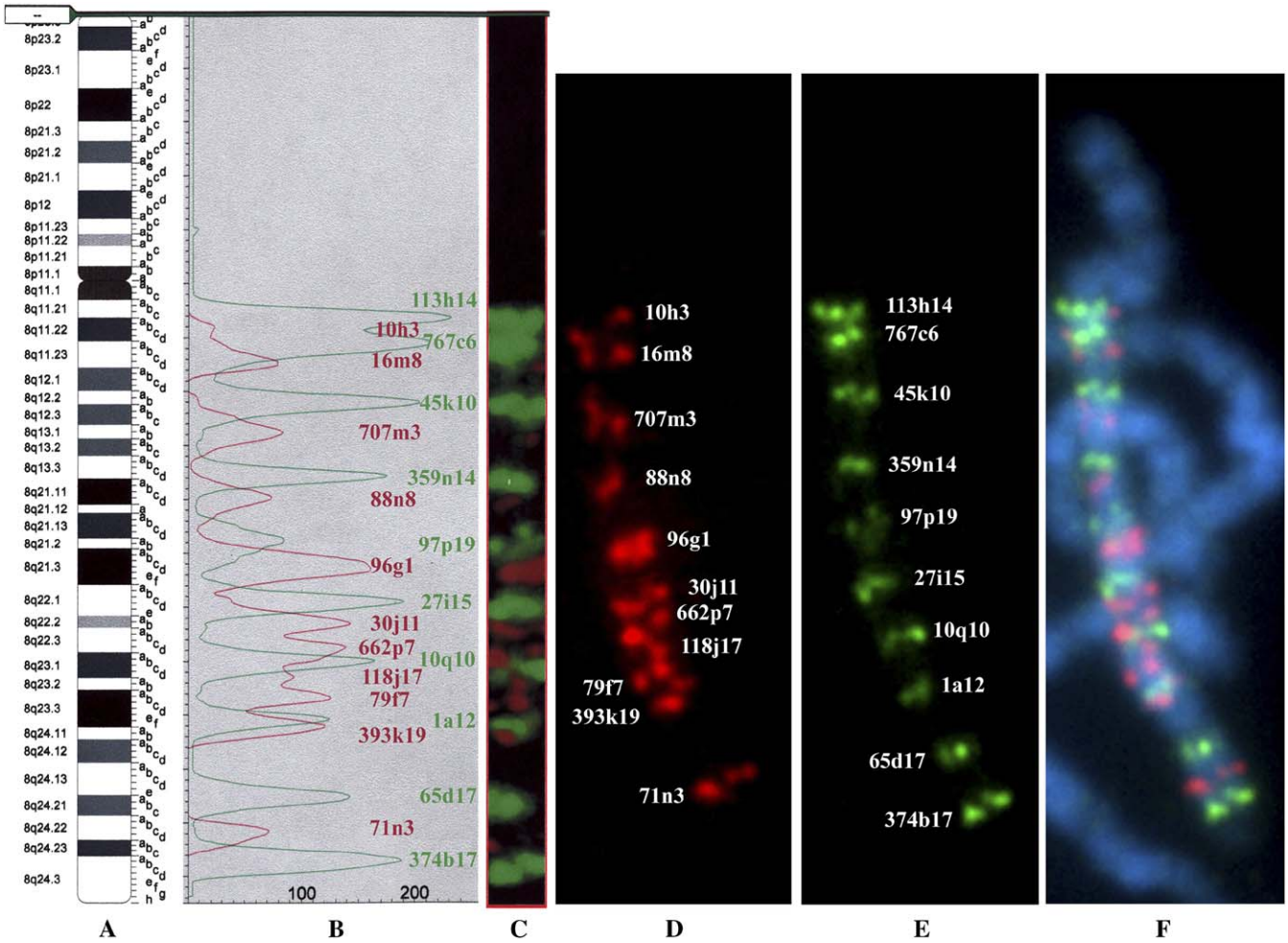


Fig. 1. High-resolution FISH mapping of two BAC clones using dual-color fluorescence. Simultaneous hybridization of 21 CCAP mapped BAC clones along the long arm of a human chromosome 8. The probes are on average separated by 2 Mb of genomic sequence. (A) An ideogram of a human chromosome 8. (B) Histogram of the signal intensities for both the fluorescein isothiocyanate (FITC) and tetramethylrhodamine isothiocyanate (TRITC) channels. (C) Images from the right three panels were imported into the CCAP software, where the chromosome is straightened. (D) 11 TRITC-labeled probes. (E) 10 FITC-labeled probes. (F) Merged image of the probes with the 4′,6-diamidino-2-phenylindole (DAPI) counterstained chromosomes.

CCAP clones with genomic base-pair coordinates. The other track, called 'NCI_fish_clone', displays CCAP clones by cytogenetic coordinates.

All the clones in the CCAP set are commercially available for purchase as purified bacterial stocks (Open BioSystems, Huntsville, AL; see Appendix). Some of the clones are available from other distributors, but those from the commercial source are precisely those used for the FISH mapping. One can purchase an individual clone as a live culture by specifying the ''NCI CCAP BAC collection,'' chromosome number, and the BAC identification number (Open BioSystems catalog number BHS 1033); to order >50 clones as a 96-well microtiter plate, specify catalog number BHS1034.

3.4. Integrating cytogenetic landmarks to the human genome

To find a human genomic region corresponding to a cytogenetic band, it is essential to annotate the sequenced human genome with such bands. The data produced from the CCAP project facilitate this annotation, because each CCAP clone has been mapped by high-resolution (850-band) FISH. The clones are also precisely localized to the sequenced human genome. Using the NCBI Map Viewer, cytogenetic data and genome sequence can be viewed side by side where the connection is made based on the CCAP clone by selecting 'NCI_FISH_clone' track and 'NCI_Clone' track.

A program, Bander, takes data for FISH-mapped clones and their location on the genome and predicts the location of cytogenetic bands on the human genome [23]. In the Bander algorithm, data from the CCAP clone set play an important role because of the high cytogenetic resolution and were therefore given higher weights than other FISH-mapped clones. Based on this program, all G-dark bands contain 13% more DNA, and all G-light bands contain 10% less DNA than was estimated based solely on the length of the cytogenetic bands and assuming that all bands

Table 1
Sequence content of CCAP clones

FISH chromosome	Clones, total no.	Insert sequenced, no.	End sequenced, no.	STS mapped, no.	Placed on the genome, no.
1	98	16	95	93	98
2	98	98	93	34	98
3	95	95	75	5	94
4	79	79	71	11	79
5	82	22	79	70	82
6	69	68	62	9	69
7	83	5	81	78	83
8	53	53	47	7	53
9	52	51	46	16	52
10	50	50	44	28	50
11	59	59	53	27	59
12	101	96	66	91	101
13	43	43	40	9	43
14	44	4	44	41	44
15	43	8	43	43	43
16	41	7	41	40	40
17	36	7	35	36	36
18	43	4	43	43	43
19	24	24	19	5	24
20	36	22	34	18	36
21	22	1	22	12	22
22	21	17	21	19	20
X	64	6	64	62	64
Y	3	3	3	1	3
Total	1,339	838	1,221	798	1,336

condense uniformly during mitosis. The prediction from Bander is being used in the NCBI Map Viewer, Ensembl Genome Browser, and UCSC human genome browser (see Appendix for URLs). Based on this annotation, navigation from cytogenetic data to the sequenced genome can occur seamlessly. Using the NCBI Map Viewer, one can browse the morbid map or the Mitelman Database's breakpoint map (see Appendix) and have corresponding sequenced genomic regions next to them.

4. Discussion

The linkage of a chromosomal position with the precisely defined sequence of the human genome will find application in almost any area of biomedical investigation, but its fullest realization to date has been in the investigation of the causes and consequences of cancer-associated chromosomal aberrations. The overarching goal of the CCAP is to facilitate and make routine what has up to now been the largely ad hoc process for the cloning and characterization of gene segments involved in cancer-associated chromosomal aberrations. The platform for achieving this goal is the presentation of a systematic integration of the cytogenetic and physical maps of the human genome. This is accomplished by the establishment of a well-characterized set of 1,339 bacterial artificial chromosome (BAC) clones spaced at 1–2 Mb intervals across the genome.

Two by two, each sequential BAC clone was localized by high-resolution (850-band stage), dual-color FISH to a specific site on human prometaphase chromosomes. For chromosome bands in which no additional cytogenetic landmarks allow for an ordering of BACs, arbitrary subdivisions of the band ('a', 'b', 'c', and so on; approximately 1 Mb) were made so that the orientation of one BAC to another could be delineated. Every BAC clone included in the CCAP set hybridizes to one and only one place in the genome and has an appreciable sequence, allowing its placement on the human genome.

The clone set is available from a commercial repository (Open Biosystems, Huntsville, AL). The clones can be visualized on the Web site supported by the National Center for Biotechnology Information (NCBI) and the National Cancer Institute (http://cgap.nci.nih.gov/Chromosomes/CCAP_BAC_Clones). From this same Web site, it is possible to link directly to the primary nucleotide sequence of the human genome, as well as a variety of other physical and conceptual databases. One can fill in the 1–2 Mb of genomic space between any two of the CCAP BAC anchor clones with additional FISH-mapped clones by looking at 'Clone' track in NCBI Map Viewer and 'FISH Clones' in the UCSC genome browser, or with additional BAC clones by choosing BAC-end clone tracks. This filling-in process is greatly facilitated by the previously reported and on-going listing of a large number of BAC clones, mapped with lower resolution and less direct connection to human genome sequence, that were generated as part of a consortium mapping enterprise [24].

The involvement of a chromosomal region defined by cytogenetic analysis and BAC clone validation can be queried for its known incidence and prevalence of association with malignant transformation. This search can now be accomplished, again, by direct linkage from the BAC clone Web sites to the Mitelman Database of Chromosome Aberrations in Cancer, or by querying the NCBI Map Viewer database that includes the 'Mitelman' track for the Mitelman database [25]. The information in the Mitelman database relates chromosomal aberrations to tumor characteristics based either on individual cases or associations. Complementary to the CCAP Database are the NCI and NCBI SKY/M-FISH and CGH Database and the Cancer Chromosomes Database (see Appendix for URLs), which are also part of the CGAP initiative [26]. The hope that the CCAP set would facilitate the cloning and characterization of cancer-associated chromosomal aberrations has been realized. For example, we ourselves used it in the process of cloning the characteristic chromosomal translocation associated with the development of mucoepidermoid carcinoma [27].

A total of 99.8% of these clones could be placed on the human genomic sequence. For most of the cases, the various sequence anchors from a clone agreed with each other when compared to available human genome sequence. For example, clone RP11-1022D13, which was FISH-mapped to 12q15c (where the 'c' denotes sublocalization within

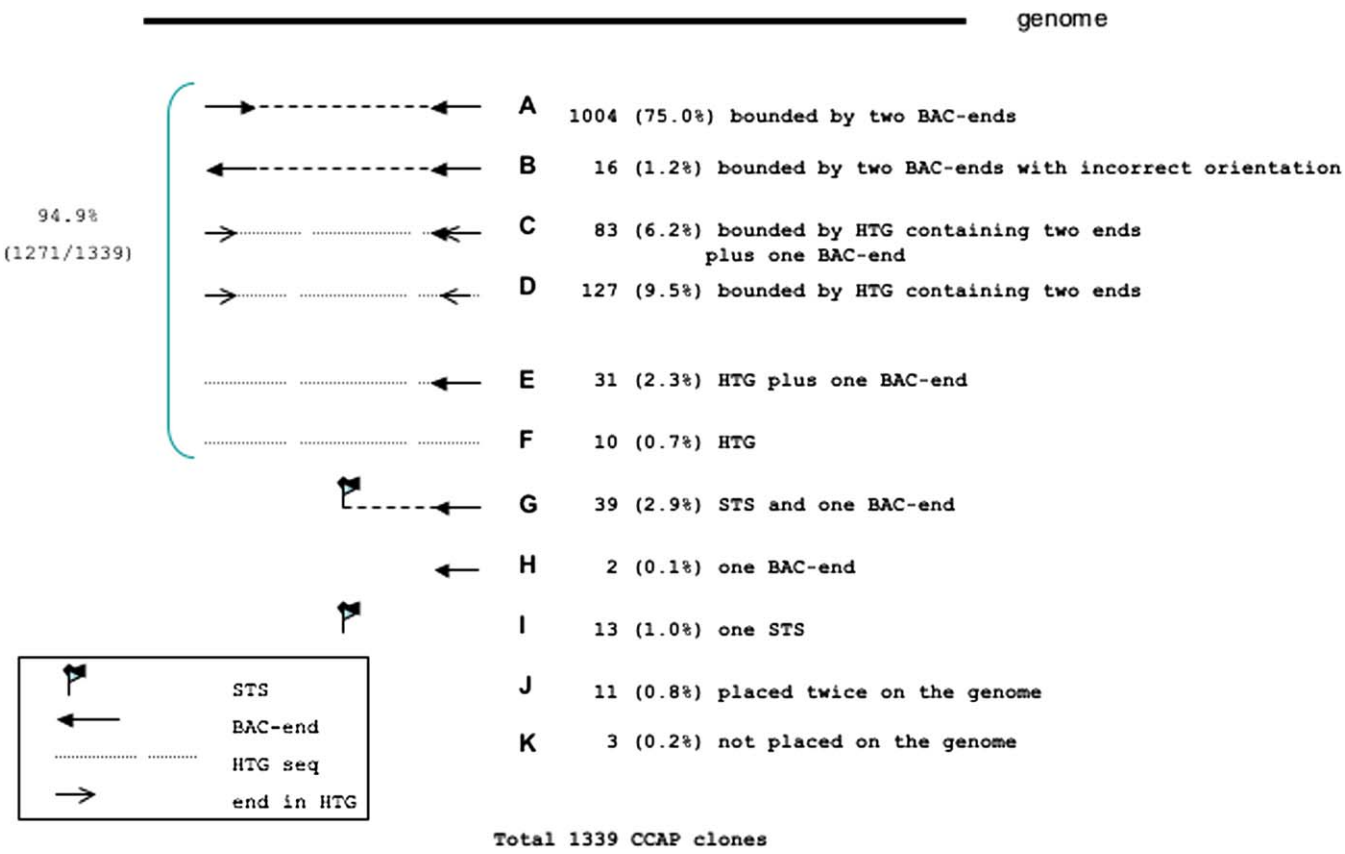


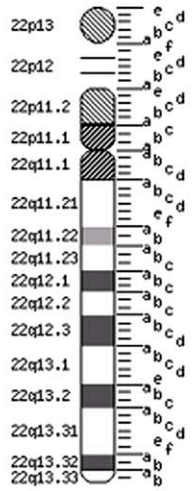
Fig. 2. Clone placement on the human genome and details of sequence anchors marking the clone boundary. (A) These clones are demarcated by two BAC-ends with correct orientation, one end in + orientation and the second end in − orientation. (B) These clones are demarcated by two BAC-ends but both BAC-ends are aligned in the same orientation. (C) These clones are demarcated by one BAC-end on one side, but they also contain HTG sequences that represent two clone ends, thereby indicating that the clone boundaries are well defined in the placement. (D) These clones are placed using HTG sequence that contains two clone ends, indicating that the clone placement represents the clone boundary well. (E) These clones are demarcated by BAC-end on one side and the boundary of the other side is unknown; however, because their placement was also supported by HTG sequences, depending on the length of the HTG sequence, some part of the clone region is represented. In some cases, this is because the second BAC-end is not aligned to the genome or aligned to different region of the genome. (F) Clones placed by using HTG sequences. (G) Clones placed by BAC-end on one side and STSs. (H) Two clones are in this category. Clones are placed by BAC-end on one side. For RP11-453D5, the second BAC-end was not aligned to the genome. For CTC-322K22, the second BAC-end was aligned to different chromosome. (I) Clones are placed by STSs. (J) Clones are placed to two locations on the genome. (K) Clones that could not be placed on the genome.

band 12q15), had the following sequence anchors: BAC end sequence CL423268.1 sequenced from the SP6 end, BAC-end sequence CL423273.1 sequenced from the T7 end, HTG phase 3 sequence AC016153.21, its previous version, HTG phase 1 sequence AC016153.20, and STS WI-17013. All these sequence anchors were placed on the genome in a region between 68,179 and 68,369 kb on chromosome 12 in NCBI human build 35.

In some cases, a clone can have sequences that disagree with FISH data and/or with other sequences. For example, clone RP11-147L21, which was FISH-mapped to 20p13f, had the following sequences: HTG phase 1 sequence AC040953.2, BAC-end sequence CL423321.1 from the SP6 end, BAC-end CL423326.1 from the T7 end, another BAC-end AQ373834.1 from the SP6 end, STS D20S103, and STS GDB:197832. All of these sequences with the exception of AQ373834.1 were placed on the genome between 376 kb and 549 kb on chromosome 20. The BAC-end sequence AQ373834.1 from the SP6 end, was placed on a position 186 Mb on chromosome 1, which is discordant with the FISH mapping data and other sequences. Because AQ373834.1 from the SP6 end is the sole sequence that identified this discordant location, this placement was ignored in the placement of clone RP11-147L21. In the CCAP Web site, AQ373834.1 was marked to indicate discordant data.

All such sequences that disagree with final clone placements are clearly marked as discordant in the CCAP Web site. There are 87 end sequences, 20 HTG sequences, and 18 STSs that fall in this category. It is likely that these inconsistencies are due to clone mix-ups, well-to-well contamination in handling bacterial cultures, or transcriptional labeling errors when the sequence data were submitted to GenBank, and therefore it is important to order CCAP clones from the CCAP-designated commercial distributor to obtain

Chromosome 22



Map Coordinate	BAC id	Placement on genome[1]	Placement summary[3]	BAC End[2]	Insert Sequence[2]	STS id[2]
22q11.1d	CTA-115F6	chr22 16Mb		B18171 T7, B18170 SP6		D22S137
22q11.21	RP11-91O6	chr22 16Mb		AQ282783 T7, AQ282780 SP6		
22q11.22a	CTA-526G4	chr22 20Mb		BZ749020 SP6, BZ749021 T7		D22S636, WI-13513+, stD22S112+
22q11.23a	CTA-322B1	chr22 22Mb		B13824 SP6, B13825 T7	Z84718+, Z84718	AL008646, D22S1106
22q11.23c	CTA-221G9	chr22 23Mb		BZ749006 T7, BZ749005 SP6, B18023 SP6	Z99916+, Z99916	AL021680
22q12.1b	CTA-992D9	chr22 25Mb		B14155 T7, B14154 SP6	AL008638	stWI-13238+
22q12.1c	CTA-57G9	chr22 27Mb		B13991 T7, B18131 T7	Z95116+, Z95116	D22S991E
22q12.2b	CTA-99F11			B14164 SP6+		
22q12.3	CTA-221H1	chr22 32Mb		BZ749008 T7, B18024 T7, BZ749007 SP6	AL008717+, AL008717	D22S629
22q12.3	CTA-415G2	chr22 31Mb		BZ749016 SP6, BZ749017 T7	Z83846+, Z83846	D22S411E
22q12.3d~13.1d	CTA-212A2	chr22 34Mb		BZ749004 T7, BZ749003 SP6	Z95114	D22S541, D22S740, WI-16714, stcN27A12T3+
22q13.1c	CTA-390B3	chr22 36Mb		B13894 SP6, B13895 T7	Z93096+, Z93096	stbK390B3.TP+

BAC-end sequence aligned in the plus orientation
BAC-end sequence aligned in the minus orientation
insert sequence
STS aligned
insert sequence has clone end information

Fig. 3. CCAP BAC clone mapping representation for part of human chromosome 22 in NCBI MapViewer. Column heading footnotes: 1. FISH-mapped clone was placed on the sequenced genome based on BAC-ends, insert sequences and STSs. 2. Sequences that are associated with the clone and used for clone placement. 4. Placement summary indicates sequences used for clone placement. *Sequences that disagree with data in 'Placement on genome' column. + Sequences that did not align to the genome or were not used for annotation.

validated clone material. These inconsistencies provide a useful and cautionary insight into interpretation of clone data for the human sequence dataset.

Eleven clones were localized to two different locations in the genome (category J in Fig. 2). For example, clone CTB-111H21 was placed to 29 Mb and also to 31 Mb region on chromosome 7. One location could not be selected over the other, because the two placements are closer than the resolution of cytogenetic bands.

For 91.3% (1,222 of 1,339) of the clones, the FISH mapping results were concordant with the in silico cytogenetic assignments predicted on the basis of the genome sequence map. A remaining 7.2% (96 of 1,339) of the clones were placed within 1 Mb of the concordant band, and 1.2% (16 of 1,339) of the clones were placed within 5 Mb of the concordant band.

Without the CCAP data and other data on FISH-mapped clones, the prediction of cytogenetic bands from genome sequence had relied on uniform conversion between base pair (bp) scale and cytogenetic bands. However, using the Bander program that relies on CCAP data and other FISH-mapped clones [23] it has been estimated that the packaging of DNA into metaphase chromosomes is uneven, with G-dark bands containing more DNA and G-light bands containing less DNA. A related finding is that for chromosomes where clones were initially chosen such that they would be separated by 1–2 Mb based on their location on the genome, there are more CCAP clones that mapped to the G-dark bands. This finding likely reflects this postulated uneven ratio of DNA packaging.

In summary, we have established a CCAP clone set of 1,339 BAC clones that have been FISH-mapped onto high-resolution chromosomes and precisely localized to the human genomic sequence at 1–2 Mb intervals. This resource facilitates systematic integration of the cytogenetic and physical maps of the human genome and will thereby help the nucleotide-level characterization of chromosomal aberrations.

Acknowledgments

Reagents for this work were supplied by Pieter de Jong, Norma Nowak, Eric Green, Raju Kucherlapati, Vivien Cheung, and Nigel Carter. None of the authors report a conflict of interest or competing interest with any commercial or academic entity mentioned in this report. We thank Dr. Michael Difilippantonio for making Fig. 1 available for use in this manuscript. This research was supported in part by the Intramural Research Program of the U.S. National Institutes of Health, National Cancer Institute, Center for Cancer Research.

Appendix

Web site URL addresses

Commonly used genome terms: http://www.ncbi.nlm.nih.gov/genome/guide/glossary.htm

The CCAP set of BAC clones [FISH-mapped BACs (CCAP)]: http://cgap.nci.nih.gov/Chromosomes/CCAP_BAC_Clones

Commercial access to the CCAP BAC clones (Open BioSystems, Huntsville, AL): http://www.openbiosystems.com

NCBI Map Viewer, *Homo sapiens* (human) genome view: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606

Ensembl Genome Browser: http://www.ensembl.org/Homo_sapiens

UCSC (University of California, Santa Cruz) human genome browser: http://www.genome.ucsc.edu/cgi-bin/hgGateway/

Mitelman Database of Chromosome Aberrations in Cancer: http://cgap.nci.nih.gov/Chromosomes/Mitelman

SKY/M-FISH and CGH Database [spectral karyotyping, multiplex FISH, and comparative genomic hybridization]: http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi

Cancer chromosomes [integrating three databases: NCI/NCBI SKY/M-FISH and CGH Database; NCI Mitelman Database of Chromosome Aberrations in Cancer; the NCI Recurrent Aberrations in Cancer]: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cancerchromosomes

NCBI Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Gene

NCBI Human Genome Resource: http://www.ncbi.nlm.nih.gov/genome/guide/human

GenBank: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide

References

[1] Strausberg RL, Dahl CA, Klausner RD. New opportunities for uncovering the molecular basis of cancer. Nat Genet 1997;15:415–6.

[2] Strausberg RL. The Cancer Genome Anatomy Project: new resources for reading the molecular signatures of cancer. J Pathol 2001;195:31–40.

[3] Kirsch IR, Green ED, Yonescu R, Strausberg R, Carter N, Bentley D, Leversha MA, Dunham I, Braden VV, Hilgenfeld E, Schuler G, Lash AE, Shen GL, Martelli M, Kuehl WM, Klausner RD, Ried T. A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. Nat Genet 2000;24:339–40.

[4] Kirsch IR, Ried T. Integration of cytogenetic data with genome maps and available probes: present status and future promise. Semin Hematol 2000;37:420–8.

[5] Mitelman F, Johansson B, Mandahl N, Mertens F. Clinical significance of cytogenetic findings in solid tumors. Cancer Genet Cytogenet 1997;95:1–8.

[6] Ried T, Heselmeyer-Haddad K, Blegen H, Schröck E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. Genes Chromosomes Cancer 1999;25:195–204.

[7] Bardi G, Johansson B, Pandis N, Bak-Jensen E, Orndal C, Heim S, Mandahl N, Andren-Sandberg A, Mitelman F. Cytogenetic aberrations in colorectal adenocarcinomas and their correlation with clinicopathologic features. Cancer 1993;71:306–14.

[8] Johansson B, Bardi G, Pandis N, Gorunova L, Backman PL, Mandahl N, Dawiskiba S, Andren-Sandberg A, Heim S, Mitelman F. Karyotypic pattern of pancreatic adenocarcinomas correlates with survival and tumour grade. Int J Cancer 1994;58:8–13.

[9] Pandis N, Idvall I, Bardi G, Jin Y, Gorunova L, Mertens F, Olsson H, Ingvar C, Beroukas K, Mitelman F, Heim S. Correlation between karyotypic pattern and clincopathologic features in 125 breast cancer cases. Int J Cancer 1996;66:191–6.

[10] Bardi G, Fenger C, Johansson B, Mitelman F, Heim S. Tumor karyotype predicts clinical outcome in colorectal cancer patients. J Clin Oncol 2004;22:2623–34.

[11] Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. Human *c-myc* oncogene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc Natl Acad Sci U S A 1982;79:7824–7.

[12] Shen-Ong GL, Keath EJ, Piccoli SP, Cole MD. Novel *myc* oncogene RNA from abortive immunoglobulin-gene recombination in mouse plasmacytomas. Cell 1982;31(2 Pt 1):443–52.

[13] Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P. Translocation of the *c-myc* gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A 1982;79:7837–41.

[14] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921 [Errata in: Nature 2001;411:720; Nature 2001;412:565].

[15] Ried T, Baldini A, Rand TC, Ward DC. Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. Proc Natl Acad Sci U S A 1992;89:1388–92.

[16] Yunis JJ, Chandler ME. High-resolution chromosome analysis in clinical medicine. Prog Clin Pathol 1978;7:267–88.

[17] Mahairas GG, Wallace JC, Smith K, Swartzell S, Holzman T, Keller A, Shaker R, Furlong J, Young J, Zhao S, Adams MD, Hood L. Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. Proc Natl Acad Sci U S A 1999;96:9739–44.

[18] Zhao S, Malek J, Mahairas G, Fu L, Nierman W, Venter JC, Adams MD. Human BAC ends quality assessment and sequence analyses. Genomics 2000;63:321–32.

[19] Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000;7:203–14.

[20] Perlin MW, Duggan DJ, Davis K, Farr JE, Findler RB, Higgins MJ, Nowak NJ, Evans GA, Qin S, Zhang J, Shows TB, James MR, Richard CW 3rd. Rapid construction of integrated maps using inner product mapping: YAC coverage of human chromosome 11. Genomics 1995;28:315–27.

[21] Schuler GD. Sequence mapping by electronic PCR. Genome Res 1997;7:541–50.
[22] ISCN 1995: an international system for human cytogenetic nomenclature (1995). In: Mitelman F, editor. Basel: S. Karger, 1995.
[23] Furey TS, Haussler D. Integration of the cytogenetic map with the draft human genome sequence. Hum Mol Genet 2003;12:1037–44.
[24] Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, Furey TS, Kim UJ, Kuo WL, Olivier M, et al. BAC Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature 2001;409:953–8.
[25] Mitelman F, Johansson B, Mertens F. Catalog of Chromosome Aberrations in Cancer. 5th ed. New York: Wiley-Liss, 1994 [last print edition; Internet edition, updated quarterly, available from http://cgap.nci.nih.gov/Chromosomes/Mitelman].
[26] Knutsen T, Gobu V, Knaus R, Padilla-Nash H, Augustus M, Strausberg RL, Kirsch IR, Sirotkin K, Ried T. The interactive online SKY/M-FISH and CGH database and the Entrez cancer chromosomes search database: linkage of chromosomal aberrations with the genome sequence. Genes Chromosomes Cancer 2005;44:52–64.
[27] Tonon G, Modi S, Wu L, Kubo A, Coxon AB, Komiya T, O'Neil K, Stover K, El-Naggar A, Griffin JD, Kirsch IR, Kaye FJ. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nat Genet 2003;33:208–13 [Erratum in: Nat Genet 2003;33:430].